Health Sciences

UM6SS EDITIONS

Volume 1 | Issue 1 Article 7

# Improved Quality: Item and Test Parameters

Chakrabartty Nath Satyendra
*Indian Ports Association, Indian Maritime University*, chakrabarttysatyendra3139@gmail.com

HEALTH SCIENCES
2020, Volume 1, ID 267
DOI: 10.15342/hs.2020.267

RESEARCH ARTICLE

# Improved Quality: Item and Test parameters

Chakrabartty Nath Satyendra
Indian Ports Association, Indian Maritime University

**ABSTRACT**

**Introduction**: Quality of a MCQ type test depends on qualities of the constituent items, assessed in terms of item reliability, item difficulty value, item discriminating value, etc. However, quality of a test involving reliability, validity, difficulty and discriminating values of the test etc. requires new approaches. Need is felt to find difficulty and discriminating values of an item and test using entire data and to derive relationships amongst them including relationship with test reliability to see impact of item deletion. **Methods:** Using angular similarity approach, measures proposed for item difficulty and item discriminating value, difficulty and discriminating value of test. Relationship derived between (i) difficulty value and discriminating value of item; (ii) difficulty value and discriminating value of a test (iii) test discriminating value and test reliability as per theoretical definition. Cronbach alpha was expressed using sum of item difficulty values and test discriminating value **Results and Discussion**: Each proposed measure ranges between 0 to 1. Discriminating value of test and item as coefficient of variation satisfy desired properties and facilitates population estimations. Intersection of item difficulty and item discriminating curves provides a data driven criterion for item deletion, impact of which on test reliability may be checked. In addition, the proposed measures facilitate testing of statistical hypothesis of departure of test reliability from unity, confidence interval of reliability, etc. Future problems suggested.

**KEYWORDS:** Difficulty value; Discriminating value; Coefficient of variation; Reliability

**Correspondence:** Pr Chakrabartty Satyendra Nath, N 304, Vivek Vihar, Sector 82, Noida – 201304, India.
Email: chakrabarttysatyendra3139@gmail.com



## INTRODUCTION

Test containing of Multiple choice questions (MCQ) are increasingly used in assessment of Medical education in various context like admission. monitoring of knowledge, path of learning, etc. The Medical council of India (MCI) conducts faculty development programs for teaching modern medical education technology (MET). (http://www.mciindia.org/information desk/for colleges/faculty development program). This includes among others multiple choice questions (MCQs) for assessment of knowledge of students [1, 2]. The quality of assessment depends on a host of factors including quality of a MCQ type test which are derived from quality of the constituent items/ Item qualities are assessed in terms of item difficulty value ($Diff_i$), item discriminating value ($Disc_i$) etc. In addition, direct measures are used to reflect quality of a test by various test statistics like test reliability ($r_{tt}$), error variance of test ($S_E^2$), test validity. However, difficulty and discriminating value of test could also indicate test qualities since major purpose of a test is to find how the test can discriminate good performers from others. Thus, discriminating value is related to the quality of the score as a measure of the trait [3].

Large number of studies have been reported on MCQ type tests administered among students of MBBS course and scores obtained by the students were used to compute $Diff_i$ and $Disc_i$ and concluded that items with good discriminating values have moderate difficult values and very difficult items may even result in negative discrimination. However, there exist gaps since none of these studies attempted to find:

i) Item discriminating value as a purely item statistic using the entire data pertaining to the item only
ii) Exact mathematical relationship between $Diff_i$ and $Disc_i$
iii) Difficulty value ($Diff_T$) and discriminating value($Disc_T$) of test and their relationship
iv) Relationship between $Diff_T$ with $Diff_i's$ and also between $Disc_T$ and $Disc_i's$
v) Data driven objective criterion for deletion of items based on behavior of $Diff_i$ and $Disc_i$
vi) Relationship between $r_{tt}$ and $Disc_T$ to see effect of item deletion on test reliability
vii) Methods of obtaining indices like $Diff_i$, $Disc_i$, $Diff_T$, $Disc_T$, etc. to facilitate population estimates of the indices and undertaking test of statistical hypothesis on equality of two similar indices.

Considering increasing use of MCQs in assessment, suggestion of [4] to find effectiveness of MCQ items, the present study aims at proposing measures of $Diff_i$, $Disc_i$, $Diff_T$, $Disc_T$ etc. without sacrificing any portion of data and making no assumption of continuous nature or linearity or normality for the observed variables or the underlying variable being measured and provide satisfactory answers to the above said gaps.

**LITERATURE SURVEY:**

Usually, item difficulty value ($Diff_i$) is computed using the entire data pertaining to the item as proportion of persons passing the item i.e. $Diff_i = \frac{k}{n} \times 100$ where $k$ denotes number of persons who gave correct answer to the item and n denotes the sample size i.e. total number of persons taking the test. Note that $0 \leq Diff_i \leq 100$ and actually reflects easiness of an item.

Discriminating value of an item ($Disc_i$) traditionally considers top 27% and bottom 27% of the data. However, out of 148 persons who took the test, [5] considered top 32% as the high group and bottom 32% as the low group and leaving aside scores obtained by the remaining persons who constituted the middle group and following Medical Education principles [6, 7] and computed $Diff_i = \frac{(H+L)}{n} \times 100$ and $Disc_i = 2 \times \frac{(H-L)}{n}$ where H and L denote respectively the number of correct responses to the item in high and low groups. [8, 9] used same formulas but considering top third as high achievers (H) and bottom third (L) as low achievers. It may be noted that an item will have negative $Disc_i$ if $L > H$. Different choices of High and Low groups have effect on $Diff_i$ and $Disc_i$. Item discrimination, without throwing any part of data are correlation between an item score and total scores on all other items on the test i.e. biserial correlation ($r_{bis}$) or point-biserial correlation ($r_{pbis}$), rank correlation (Spearman's $\rho$), etc [10]. Index of intertertile discriminating power using *t*- statistic for testing equality of mean scores was suggested [11]. However, *t*- statistic requires normal distribution of scores and independent samples.

Different measures of $Disc_i$ resulted in contrasting relationship between $Diff_i$ and $Disc_i$. Correlation between $Diff_i$ and $Disc_i$ was 0.563 [12] and as low as 0.191 [9]. A dome-shaped curve was observed between them i.e. non-linear relationship [4].

**PROPOSED METHODS:**

Consider a test containing $m$-MCQ items (1 for correct answer and 0 otherwise) administered to $n$-respondents. Mean and variance of test score can also be looked from angular association approach described below:

Let X= $(X_1, X_2, \ldots, X_n)^T$ be the test score vector. Let the maximum possible test score vector be I of order $n \times 1$ where each component is $m$. If the two vectors X and I are same, the test is extremely easy since each subject has got maximum possible score and the test fails to discriminate the subjects. Let the angle between the vectors X and I is $\theta$ which can be computed by $Cos\theta = \frac{X^T I}{\|X\|\|I\|}$ where $\|X\|= \sqrt{\sum_{i=1}^{n} X_i^2}$ is the length of the vector X. $\|I\|$ is defined accordingly. Here, $\|I\| = m\sqrt{n}$ .

Note that $Cos\theta = \frac{m\sum X_i}{\|X\|\, m\sqrt{n}} \Longrightarrow \bar{X} = \frac{\|X\|Cos\theta}{\sqrt{n}}$ (1)

and $Sin^2\theta = 1 - \frac{n.\bar{X}^2}{\|X\|^2} \Longrightarrow \|X\|^2 Sin^2\theta = \|X\|^2 - n.\bar{X}^2$

$\Longrightarrow$Test Variance $S_X{}^2 = \frac{\|X\|^2 Sin^2\theta}{n}$ (2)

Similarly, score vector $X_i$ of the *i*-th item contains zeros and ones. Let $I_i$ be the maximum possible score vector for an item where each component is equal to 1. If *k*–persons ($k \leq n$) answer the *i*-th item correctly, $\|X_i\| = \sqrt{k}$ , $\|I_i\| = \sqrt{n}$ and $\cos\theta_i = \sqrt{\frac{k}{n}}$

Difficulty value of test is proposed to be defined as

$$Diff_T = \frac{\|X\|}{\|I\|}\cos\theta = \frac{\bar{X}}{m} \quad (3)$$

Clearly, $0 \leq Diff_T \leq 1$ and Higher value of $Diff_T$ implies easier is the test.

Difficulty value of an item can be defined as

$$Diff_i = Cos^2\,\theta_i = \frac{k}{n} \quad (4)$$

Note that, $0 \leq Diff_i \leq 1$.

As *k* increase, $Diff_i$ also increases. Thus, the $Diff_i$ curve is positively slopped

Equation (4) agrees with the usual idea of item difficulty value as proportion of persons passing an item. However, the approach also helps us to express $Diff_T$ as a function of $Diff_i's$.

Note that $\bar{X} = \frac{\sum_{i=1}^{m} k_i}{n} = \sum_{i=1}^{m} Diff_i$.

Thus, from (3), $Diff_T = \frac{\sum_{i=1}^{m} Diff_i}{m}$ (5)

Equation (5) gives relationship between $Diff_T$ and $Diff_i's$

Discriminating value of test:

Discriminating value of a test is a measure of dissimilarity between the vectors X and I and can be given as $Disc_T = tan\theta = \frac{S_X}{\bar{X}}$ [From (1) and (2)] (6)

Thus, $Disc_T$ is equal to coefficient of variation (CV) of the test scores.

Discriminating value of item:

Item discriminating value can be similarly defined by $Disc_i = \frac{S_{X_i}}{\bar{X_i}}$ (7)

where $\bar{X_i}$ and $S_{X_i}$ denote respectively mean and SD of the *i*-th item.

Score of the *i*–th item follows a Binomial distribution with parameters $n$ and $p_i$ (probability of correct answer) where mean and SD are $np_i$ and $\sqrt{np_i q_i}$ respectively, where $q_i = 1 - p_i = \frac{n-k_i}{n}$. Thus, co-efficient of variation of the *i*-th item ($CV_i$) is $\frac{\sqrt{np_i q_i}}{np_i} = \frac{1}{\sqrt{n}}\sqrt{\frac{q_i}{p_i}}$

So, $Disc_i = \frac{S_{X_i}}{\bar{X_i}} = \sqrt{\frac{n-k_i}{nk_i}}$ (8)

Clearly, $Disc_i$ by (8) is truly an item parameter and avoids performance of persons in all other items of the test and $0 \leq Disc_i < 1$. The equation (8) considers ratio of number of success and failures in the *i*-th item and avoids usual range of item discrimination between (-) 1.0 to (+) 1.0 and thus avoids difficulty in interpretation of negative discriminating values. Since $Disc_i$ decreases with increase in $k$, $Disc_i$ curve is negatively sloped and negatively correlated with $Diff_i$.

The percentage discriminating value of the $i$-th item = $100.Disc_i$ (9)

Relationship between $Disc_i$ and $Diff_i$ :

From (8) and (4), we get $Disc_i^{\,2} = \frac{1-Diff_i}{n.Diff_i} = \frac{1-Diff_i}{k}$ (10)

i.e. square of discriminating value of an item is equal to (1-difficulty value of the item) divided by number of correct response ($k$) to the item. .

Equation (10) shows the non-linear relationship between $Diff_i$ and $Disc_i$. While low value of $k$ implies higher $Disc_i$, a higher $Diff_i$ implies lower $Disc_i$ Thus, the correlation between $Diff_i$ and $Disc_i$ is negative. If $k_0$ is a solution of equation (10) then at $k_0$, $Disc_i = Diff_i$ i.e. the point of intersection of the positively slopped $Diff_i$ curve and negatively slopped $Disc_i$ curve. In general, $k_0$ will be a non-integer. In that case, the nearest integer value of $k_0$ may be taken. Item with scores lying beyond a small neighborhood of $k_0$ may be taken as a data driven criterion for item deletion.

Note that: $k_i = 0 \Rightarrow$ no subjects could pass the item $\Rightarrow$ $Diff_i = 0$ and $Disc_i$ is undefined for the item. Similarly, $k_i = n \Rightarrow$ all the subjects pass the item $\Rightarrow$ $Diff_i = 1$ and $Disc_i$ =0. Such items to be deleted without further investigation.

Relationship between $Diff_T$ and $Disc_T$:

From (3) and (6), we get $Diff_T.Disc_T = \frac{S_X}{m}$ (11)

Product of $Diff_T$ and $Disc_T$ increases with increase in test variance and also with decrease of length of the text.

Relationship between test reliability and $Disc_T$:

Variance of the $i$-th item $S_{X_i}^2 = \bar{X_i}^2 . Disc_i^2$ [from (7)] $\Rightarrow$ $\sum_{i=1}^{m} S_{X_i}^2 = \sum_{i=1}^{m} \bar{X_i}^2 . Disc_i^2$

Test variance $S_X^2 = \bar{X}^2 .\ Disc_T^{\,2}$ [from (6)]. Thus, test reliability in terms of Cronbach alpha is

$$\alpha = (\frac{m}{m-1})(1 - \frac{\sum_{i=1}^{m} \bar{X_i}^2 . Disc_i^2}{\bar{X}^2 . Disc_T^{\,2}}) \quad (12)$$

Equation (12) expresses $\alpha$ in terms of $Disc_T$ and $Disc_i's$. Impact of deletion of an item on alpha can be worked out using (12) keeping in mind that Cronbach alpha assumes uni-dimensinality of test. However, considering theoretical definition of test reliability, $r_{tt} = \frac{S_T^2}{S_X^2}$, we get

$$r_{tt}(Disc_T)^2 = \frac{S_T^2}{\bar{X}^2} = (\frac{S_T}{\bar{X}})^2 = (\frac{S_T}{\bar{T}})^2 \quad (13)$$

i.e. product of test reliability and square of test discriminating value is equal to square of CV of true scores. Both (12) and (13) indicate that $Disc_T$ has a negative non-linear relationship with $r_{tt}$

Equation (13) can be verified by computing $r_{tt}$ as per theoretical definition, by the method given by [13] involving dichotomization of the test in $g$-th and $h$-th subtests where the subtests are parallel. This helps in computation of error variance ($S_E^2$) and true score variance ($S_T^{\,2} = S_X^{\,2} - S_E^2$) from single administration of a test with binary MCQ items by the following:

$$r_{tt} = 1 - \frac{\|X_g\|^2 + \|X_h\|^2 - 2\,\|X_g\|\|X_h\| cos\theta_{gh}}{n\, S_X^2} \quad (14)$$

$$S_E^2 = \frac{1}{n}[\,\|X_g\|^2 + \|X_h\|^2 - 2\,\|X_g\|\|X_h\| Cos\,\theta_{gh}\,] \quad (15)$$

**ESTIMATION AND TESTING OF HYPOTHESIS**

Sample $Diff_i$ is a proportion [from (4)] and sample $Diff_T$ is the average of the $Diff_i's$ [from (5)]. Statistical inferences about proportion and average are given in any text-book on statistical inferences.

Each of $Disc_T$ and $Disc_i$ is a coefficient of variation (CV) [from (6) and (7)]. Unbiased estimator of population CV for data following normal distribution and separately for log-normal distribution has been proposed respectively by [14] and [15]. McKay's Chi-square approximation of CV is often used for statistical inferences for normally distributed data [16].

Obtained values of sample $S_E^2$ , $S_T^2$ may help to estimate $\sigma_E^2$ , $\sigma_T^2$ for the population and use them directly to find estimate of population reliability and confidence interval of true score of a student given his/her observed score.

Reliability as per equation (14) also helps to test whether the population reliability is equal to one i.e. to test $r_{tt} = 1$. Since $r_{tt} = \frac{S_T^2}{S_X^2}$ as per the definition, the test is equivalent to testing $H_0: \sigma_X^2 = \sigma_T^2$ which can be tested using usual $F = \frac{S_X^2}{S_T^2}$ and reject $H_0$ if the test statistic $F$ is too large i.e. if $F > F_{\frac{\alpha}{2}(N-1,N-1)}$. Equivalently, the null hypothesis $H_0$ : $\sigma_E^2$ is equal to a small positive number say 0.01 can be tested using $\chi^2$ test of variance and reject $H_0$ if $T = (N-1)(\frac{SE}{0.01}^2) > \chi^2_{1-\alpha,N-1}$ [17].

Formula for testing $H_0 : \sigma_E^2 = 0.01$ can easily be converted to an interval estimation of $\sigma_E^2$ as $\sqrt{\frac{(N-1)S_E^2}{\chi^2_{(1-\frac{\alpha}{2}),(N-1)}}} \le \sigma_E \le \sqrt{\frac{(N-1)S_E^2}{\chi^2_{(\frac{\alpha}{2}),(N-1)}}}$ and confidence interval of population reliability $r_{tt} = \frac{\sigma_T^2}{\sigma_X^2}$ can be found as $\frac{S_T^2/S_X^2}{F_{\frac{\alpha}{2}}} \le \frac{\sigma_T^2}{\sigma_X^2} \le \frac{S_T^2/S_X^2}{F_{1-\frac{\alpha}{2}}}$

Empirical illustration:

Real life data on MCQ type test with 50 items ($m$), 911 persons ($n$) resulted in:

1.$Diff_T = 0.40990$; $Disc_T$= 0.16872. Thus, the test was moderately difficult and had rather poor discriminating power.

2. Correlation between $Diff_i$ and $Disc_i$ was (-) 0.57958.

3. Positively slopped item difficulty curve and negatively slopped item discriminating curve intersected at $k_0$= 368 (nearest integer). For $k$ = 368, item difficulty was 40.39% and tem discriminating was 40.24%.

4. Deletion of items:

The items to be ignored could be those lying outside the interval ($k_0 \pm$ 2SD of item scores or ($Diff_{i\ at\ k=k_0} \pm$ 2SD of item difficulty scores) or ($Diff_{i\ at\ k=k_0} \pm$ 2SD of item difficulty scores item discriminating scores). Mean, SD and acceptance regions are shown in Table 1.

**Table 1: Mean, SD and acceptance region**

| Description | Mean | SD | Acceptance region |
|---|---|---|---|
| Item scores | 20.49 | 167.59 | 368±335.18 ($k_0$ ± 2SD of item scores) |
| Item difficulty values | 0.02249 | 0.1839 | 0.02249 ±0.3679 ($Diff_{i\ at\ k=k_0}$ ± 2SD of $Diff_i$) |
| Item discriminating values | 0.00258 | 0.02544 | 0.00258 ±0.05089 ($Disc_{i\ at\ k=k_0}$ ± 2SD of $Disc_i$) |

Each of the above method resulted in discarding the item with $k$=30, being extremely difficult i.e. lowest Diff. value (0.033) and highest Disc. value (0.17954).

Deletion of an item will change values of $Diff_T$ and $Disc_T$. For example, if the most difficult item with $k$=30 is deleted, new $Diff_T$ increased to 0.417595 from original value of 0.40990 and new $Disc_T$ got reduced to 0.17392 from original value of 0.16872.

The data had 16 items with $k$-values less than 368 ($k_0$) and 34 items with $k$-values exceeding $k_0$ (rather easy items). Easy items with high $k$-values (i.e. high Diff. values say $\geq 0.70$ implying low Disc. values$\leq 0.022$) may also be considered for discarding. Adoption of this criteria implies discarding of additional four items (viz. items with $k$= 672; 670; 654 and 645).

However, deletion of items is advisable only when reliability of the test improves upon deletion

5. Splitting the test by the iterative process resulted in $\overline{X_g} = \overline{X_h}$ = 10.25 and $|S_g - S_h|$ = 0.418. Marginal difference (0.418) between the SDs of the $g$-th and $h$-th tests (much less than the same obtained from odd–even split half). Accordingly, splitting half as per the iterative process was considered better for almost equality of means and SDs.

Here, $\|X_g\| = 315.6169$; $\|X_h\| = 315.6058$ and $Cos\theta_{gh} = 0.975479$

Theoretical reliability of the test $r_{tt} = 1 - \frac{\|X_g\|^2 + \|X_h\|^2 - 2\|X_g\|\|X_h\|cos\theta_{gh}}{n.S_X^2}$ = 0.551577 and

$S_E^2 = \frac{1}{n}[\|X_g\|^2 + \|X_h\|^2 - 2\|X_g\|\|X_h\|Cos\theta_{gh}]$ = 5.362239

and true score variance = 6.595749

Theoretical reliability of the test was lower than Cronbach $\alpha$ (0.78)

## CONCLUSION

New measures of difficulty and discriminating values of MCQ items and test were proposed considering angular similarity. The measures considered entire data and not only top 27% and bottom 27% of data. Difficulty value of a test ($Diff_T$) is defined keeping harmony with the usual notion of difficulty value of a test which actually measures degree of easiness of a test. Discriminating value of a test ($Disc_T$) is the ratio of SD and mean of the test score. Similarly, discriminating value of an item ($Disc_i$) is equal to the ratio of SD and mean of the item score i.e. coefficient of variation (CV). Here, $0 \leq Disc_T \leq 1$ and similar inequalities hold for $Disc_T$, $Diff_T$ and $Diff_i$ . Discriminating value of test and also item in terms of CV has desired properties and facilitates population estimation.

Relationship derived between (i) item difficulty value and item discriminating value; (ii) difficulty value and discriminating value of a test (iii) test discriminating value and test reliability as per theoretical definition. Cronbach alpha was expressed and computed using sum of item difficulty values and test discriminating value.

As number of correct answer to an item ($k$) increases, item difficulty curve increases and item discriminating curve decreases. The point of intersection of the two curves ($k_0$)is a data driven criterion which may also be considered in deciding the items to be deleted which are lying outside acceptance region defined as an interval ($k_0 \pm \Delta$) where Δ could be taken as 2SD of distribution of Item scores or $Diff_i$ $or$ $Disc_i$. However, actual deletion of items needs to consider impact of such deletions on reliability of the test.

The proposed measures and their relationship with theoretically defined reliability also helps in testing statistical hypothesis of $r_{tt} = 1$; $\sigma_E^2 = \epsilon$ where $\epsilon$ is a small positive number; confidence interval of population reliability, etc.

Future investigations may be undertaken to investigate the proposed measures and their factors with multiple data sets.

## REFERENCES


[1] Epstein, R. M. Medical education—Assessment in medical education. The New England Journal of Medicine, 2007; 356(4), 387–396.

[2] Nedeau-Cayo, R., Laughlin, D., Rus, L., & Hall, J. Assessment of item-writing flaws in multiple-choice questions. Journal for Nurses in Professional Development, 2013; 29, 52–57.

[3] McDonald, R.P. Test theory: A unified treatment. Mahwah (NJ): 1999; Lawrence Earlbaum Associates, Inc.

[4] Sim, S.M., & Rasiah, R. I. Relationship between item difficulty and discrimination indices in true/False Type multiple choice questions of a Para-clinical multidisciplinary paper. Annals of the Academy of Medicine, Singapore, 2006, 35(2), 67–71.

[5] Gajjar Sanju, Sharma Rashmi, Kumar Pradeep, Rana Manish. Item and test analysis to identify quality multiple choice questions (MCQS) from an assessment of medical students of Ahmedabad, Gujarat, Indian J. Community Med. 2014, 39; 17 – 20.

[6] Ananthkrishnan N. The item analysis. Medical Education Principles and Practice. 2nd ed. Pondicherry: JIPMER; 2000. p. 131-137.

[7] Tejinder S, Piyush G, Daljit S. Principles of Medical Education. 3rd ed. 2009; India: Jaypee Brothers, Medical Publishers Pvt. Limited; .p. 70-77

[8] Shete, A.N., Kausar, A.,Lakhkar,K. and Khanm S. T. Item analysis: An evaluation of Mcqs in physiology examination, J Contemp Med Edu., 2014; 3(3); 106-109, DOI: 10.5455/jcme.20151011041414

[9] Pande. S.S, Santosh R. Pande, S.R., Parate, V.R., Nikam, A.P. and Agrekar, S.H. Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology, South-East Asian Journal of Medical Education, 2013; 7 (1), 45 - 50

[10] Tzuriel, D. and M. Samuels. Dynamic assessment of learning potential: Inter-rater reliability of deficient cognitive functions, type of mediation and non-intellective factors. Journal of Cognitive Education and Psychology, 2000; 1: 41-64.

[11] Preston, Carolyn C. and Colman, Andrew M.: Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychologica, 2000; 104; 1-15

[12] Rao, C., Kishan Prasad, H. L., Sajitha, K., Permi, H., & Shetty, J.: Item analysis of multiple choice questions: Assessing an assessment tool in medical students. International Journal of Education and Psychological Research, 2016; 2, 201–204.

[13] Chakrabartty, Satyendra Nath (2020): Reliability of Test Battery, Methodological Innovations. 2020; May-August; 1–8. DOI: 10.1177/2059799120918340

[14] Sokal, R.R. and Rohlf, F.J.: Biometry: The Principles and Practice of Statistics in Biological Research (3rd Ed.), 1995, New York: Freeman

[15] Julious, S.A. and Debarnot, C.A. :Why are Pharmacokinetic Data Summarizes by Arithmetic Mean? Journal of Bio-pharmaceutical Statistics, 2000, 10(1), 55-71

[16] Forkman, J: Estimator and tests for common co-efficient of variation in Normal Distribution. Communications in Statistics – Theory & Methods, 2013; 38 (2); 233-251

[17] Snedecor, G.W. and Cochran, W.G.: Statistical Methods. 6th Edition, 1983, Oxford and IBH, New Delhi.